Allele-specific quantitative proteomics unravels molecular mechanisms modulated by cis-regulatory PPARG locus variation

Heekyoung Lee, Kun Qian, Christine von Toerne, Lena Hoerburger, Melina Claussnitzer, Christoph Hoffmann, Viktoria Glunk, Simone Wahl, Michaela Breier, Franziska Eck, Leili Jafari, Sophie Molnos, Harald Grallert, Ingrid Dahlman, Peter Arner, Cornelia Brunner, Hans Hauner, Stefanie M. Hauck, and Helmut Laumen

1Else Kroener-Fresenius-Center for Nutritional Medicine, Chair of Nutritional Medicine, Technische Universität München, 85354 Freising-Weihenstephan, Germany, 2ZIEL - Institute for Food & Health, Technische Universität München, 85354 Freising-Weihenstephan, Germany, 3Clinical Cooperation Group Nutrigenomics and Type 2 Diabetes, Helmholtz Zentrum München and Technische Universität München, 85354 Freising-Weihenstephan, Germany, 4German Center for Diabetes Research (DZD), Germany, 5Research Unit Protein Science, Helmholtz Zentrum München, 85764 Neuherberg, Germany, 6Else Kroener-Fresenius-Center for Nutritional Medicine, Paediatric Nutritional Medicine, Technische Universität München, 85354 Freising-Weihenstephan, Germany, 7Hebrew SeniorLife Institute for Aging Research, Harvard Medical School, Boston, MA 02131, USA, 8Else Kroener-Fresenius-Center for Nutritional Medicine, Chair of Molecular Nutritional Medicine, Technische Universität München, 85354 Freising-Weihenstephan, Germany, 9Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, 85764 Neuherberg, Germany, 10Institute of Epidemiology II, Helmholtz Zentrum München, 85764 Neuherberg, Germany, 11Department of Medicine, Huddinge, Karolinska Institutet, SE-141 86 Stockholm, Sweden, 12Klinik für Hals-Nasen-Ohrenheilkunde, Universitätsklinik Ulm, 89075 Ulm, Germany, 13Else Kroener-Fresenius-Center for Nutritional Medicine, Klinikum rechts der Isar, Technische Universität München, 81675 Munich, Germany and 14Institute of Experimental Genetics, Helmholtz Zentrum München, 85764 Neuherberg, Germany

ABSTRACT

Genome-wide association studies identified numerous disease risk loci. Delineating molecular mechanisms influenced by cis-regulatory variants is essential to understand gene regulation and ultimately disease pathophysiology. Combining bioinformatics and public domain chromatin information with quantitative proteomics supports prediction of cis-regulatory variants and enabled identification of allele-dependent binding of both, transcription factors and coregulators at the type 2 diabetes-associated PPARG locus. We found rs7647481A nonrisk allele binding of Yin Yang 1 (YY1), confirmed by allele-specific chromatin immunoprecipitation in primary adipocytes. Quantitative proteomics also found the coregulator RING1 and YY1 binding protein (RYBP) whose mRNA levels correlate with improved insulin sensitivity in primary adipose cells carrying the rs7647481A nonrisk allele. Our findings support a concept with diverse cis-regulatory variants contributing to disease pathophysiology at one locus. Proteome-wide identification of both, transcription factors and coregulators, can profoundly improve understanding of mechanisms underlying genetic associations.

INTRODUCTION

Genome-wide association studies (GWAS) identified thousands of loci associated with diverse diseases (1). The majority of variants are located in noncoding DNA regions and have been proposed to affect transcriptional regulation. Advances of the ENCODE project and novel bioinformatics approaches improved the identification of cis-regulatory
variants at complex loci (2–7). Moreover, deciphering allele-specific binding of transcription factors is essential to unravel the mechanisms ultimately affecting gene expression (6,8–10). However, the identification of allele-specific binding of transcription factors and coregulators remains elusive in most cases, despite the well-established importance of a coordinated interaction between transcription factors, coregulators, and the basal transcriptional machinery for regulation of gene expression (11). Thus, in most cases the precise molecular mechanisms underlying associations between variants and disease risk remain unknown.

Quantitative protein–DNA proteomics, coupling affinity chromatography with liquid–chromatography tandem mass spectrometry (LC–MS/MS) was reported for identification of enhancer-binding proteins and for allele-specific DNA binding transcription factors (6,12,13). In this study, we introduce label-free quantitative proteomics on salt eluted sub-fractions containing native complexes, circumventing limitations of metabolic labeling strategies such as time-consuming workflows, high-costs or inefficiencies (14), or artifacts (15). The preserved allele-specific DNA-protein binding activity in the eluted sub-fractions significantly reduced complexity and allowed high coverage of quantified proteins and thereby identification of both, transcription factors and coregulators (Figure 1A).

The Peroxisome proliferator-activated receptor gamma (PPARG) locus has been robustly associated with type 2 diabetes (16–18). PPARG is crucial for adipogenesis, lipid metabolism, and systemic insulin sensitivity (19, 20) with the PPARG2 isoform being mainly expressed in adipocytes (21–24). In a previous study, we reported that adipose tissue PPARG locus expression quantitative trait locus (eQTL) data revealed allele-specific regulation of PPARG expression (6). In the same study, using bioinformatics phylogenetic module complexity analysis (PMCA) we found the regulatory variant rs4684847. This variant overlaps with a homeobox transcription factor binding site (TFBS), a common feature inferred for type 2 diabetes risk variants. We further reported rs4684847C risk allele binding of the Paired Related Homeobox 1 (PRRX1) transcription factor. PRRX1 represses PPARG expression, and negative correlation of PRRX1 mRNA levels with insulin-sensitivity supports contribution to insulin resistance phenotype at the PPARG locus (6).

In the current study, combining PMCA (6) with public domain DNase sequencing (DNase-seq) and chromatin immunoprecipitation sequencing (ChIP-seq) data we inferred rs7647481, a second PPARG locus cis-regulatory variant. We show that label-free proteomics can find both, transcriptional and protein–DNA interactions at the PPARG locus during adipogenic differentiation of primary human adipocyte stem cells (36), GSE21366, genomic coordinates are given conform to hg19. (D) Localization of cis-regulatory (red) and non cis-regulatory (grey) variants subjected to workflow (A2–6) relative to transcriptional start site of the PPARG1–3 mRNA isoforms. rs7647481 overlapping with both, day 3 and day 9, tested late stage of adipogenesis histone modification regions (Figure 1C) and with adipocyte DNase-seq regions (see Supplementary Figure S1). * rs4684847 previously identified as specifically overlapping with homeobox TFBS (6). Blue boxes = coding exons, dashed white boxes = untranslated exons, blue lines = introns, black arrows = promoters.

---

**Figure 1.** Discovery of allele-specific binding proteins at cis-regulatory variants. (A) Workflow: (1) cis-regulatory variant prediction at disease associated variants (PPARG) in high LD ($r^2 \geq 0.7$) (6) by integrating bioinformatics phylogenetic TFBS module complexity analysis and regulatory chromatin marks; (2) protein–DNA binding assessed by Cy5 labeled oligonucleotides matching the risk and nonrisk allele, respectively, in electrophoretic mobility shift assay (EMSA); (3) protein enrichment with biotin (bio) labeled oligonucleotides on streptavidin-beads (str) and elution of native protein complexes with increasing concentration of NaCl; (4) protein–DNA binding in eluted fractions; (5) protein identification and quantification by LC–MS/MS and subsequent label-free quantitative analysis; and (6) molecular mechanisms, experimental and genetics verification of significant allele-specific binding transcription factors and related coregulators. (B–D) Bioinformatics and public domain epigenomic marks of regulatory regions infer the cis-regulatory variant rs7647481 at the PPARG locus (related to Supplementary Figure S1). (B) PMCA analysis of cross-species TFBS pattern conservation predicted six indicated candidate cis-regulatory SNPs at complex regions (6) (red) out of 23 noncoding proxy SNPs ($r^2 > 0.7$) at the type 2 diabetes (T2D) associated PPARG locus (tagSNP rs1801282). (C) Overlap of six variants identified in (B) with H3K27ac (histone H3–lysine 27 acetylation), H3K4me1 and H3K4me2 (histone H3–lysine 4 mono- and di-methylation) histone modification regions at the PPARG locus during adipogenic differentiation of primary human adipocyte stem cells (36), GSE21366, genomic coordinates are given conform to hg19. (D) Localization of cis-regulatory (red) and non cis-regulatory (grey) variants subjected to workflow (A2–6) relative to transcriptional start site of the PPARG1–3 mRNA isoforms. rs7647481 overlapping with both, day 3 and day 9, tested late stage of adipogenesis histone modification regions (Figure 1C) and with adipocyte DNase-seq regions (see Supplementary Figure S1). * rs4684847 previously identified as specifically overlapping with homeobox TFBS (6). Blue boxes = coding exons, dashed white boxes = untranslated exons, blue lines = introns, black arrows = promoters.
tion factors and cofactors at the rs7647481A nonrisk and rs4684847C risk allele, supporting the prediction of both regulatory variants. For rs7647481, we present novel experimental and human genetics data supporting the pathophysiological relevance of the transcription factor YY1 and cofactor RYBP.

MATERIALS AND METHODS

Cell culture

The SGBS human preadipocytes cell strain, primary human preadipocytes, Huh7 human hepatoma, C2C12 mouse myoblast, INS-1 rat insulinoma, 293T human embryonic kidney and the HIB1B mouse brown adipocyte cell line were cultured, induced and differentiated as previously described (6,25). Ethical committee approval for primary human preadipocytes was obtained from the Faculty of Medicine of the Technical University of Munich, Germany.

Preparation of nuclear extracts

Nuclear protein extracts from human SGBS preadipocytes, differentiated SGBS adipocytes, and human primary preadipocytes were prepared as described (26). Briefly, cells were washed with phosphate-buffered saline (PBS), lysed in pre-chilled buffer A (10 mM HEPES pH 8.0, 10 mM KCl, 0.1 mM EDTA, 1 mM DTT, protease- and phosphatase inhibitor mix (Roche, Mannheim, Germany), 0.6% Nonidet-P40). Nuclei were collected by centrifugation, lysed in buffer B (20 mM HEPES pH 7.9, 400 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 1 mM PMSF, 20% v/v glycerol, protease and phosphatase inhibitors (Roche, Mannheim, Germany), incubated on ice for 15 min, and the nuclear extract supernatant collected for further analysis. Protein concentration was quantitated using Bradford assay. HIB1B nuclear extracts were prepared for further analysis. Protein concentration was quantitated using Bradford assay. HIB1B nuclear extracts were prepared for further analysis.

EMSA

Non-radioactive EMSA was performed with Cy5-labeled oligonucleotide as previously described (6) with some modifications. 5 μg of nuclear protein and 0.35 μg of poly (dI-dC) (Roche, Mannheim, Germany) were used for each reaction. The reaction conditions for affinity chromatography are in general similar to those used for EMSA. To optimize DNA-protein binding in EMAS, several conditions were tested, such as amount of proteins, concentration of the nonspecific competitor poly (dI-dC) and salt concentration in the gel binding buffer (3% (v/v) glycerol, 0.7 mM MgCl₂, 0.4 mM EDTA, 0.4 mM DTT, 37 mM NaCl, 0.7 mM Tris–HCl pH 7.5). In competition EMSA, 33-fold molar excess of unlabeled allele-specific probes and probes with perfect binding sites for YY1, MyoD, CdxA and a scrambled oligonucleotide were added to the reaction, prior to addition of Cy5-labeled probes. Binding reactions were incubated for 20 min at 4°C. In supershift experiments, nuclear extracts were pre-incubated with 0.8 μg of anti-YY1 (sc-281x, Santa Cruz Biotechnology, Heidelberg, Germany) or isotype control IgG (sc-2027, Santa Cruz Biotechnology, Heidelberg, Germany) antibodies, respectively, for 30 min at 4°C. All EMSA experiments were replicated at least three times.

Affinity chromatography using magnetic beads

To isolate and enrich allele-specific binding proteins, we performed magnetic beads-based affinity chromatography. According to the manufacturer’s instructions, streptavidin coupled magnetic beads (Dynabeads M-280, Invitrogen, Darmstadt, Germany) were washed and collected using Binding & Wash buffer (Dynabeads M-280, Invitrogen, Darmstadt, Germany) and Magnetic particle separator (MagnaSep™, Invitrogen, Darmstadt, Germany), discarding the supernatant. Magnetic beads were coupled to biotinylated oligonucleotides at 4°C overnight. We tested different conditions for optimal binding of beads to oligonucleotides and found that both the concentration of oligonucleotides and beads, as well as incubation time were critical for the coupling efficiency of oligonucleotides to beads. To prevent undesired reaction with streptavidin, the magnetic beads were incubated with the Binding & Wash buffer containing 8.2 μM biotin for 1 h at room temperature. The magnetic beads were then washed with wash buffer followed by equilibration with 1× binding buffer (10 mM Tris–HCl, 1 mM MgCl₂, 0.5 mM EDTA, 0.5 mM DTT, 4% v/v glycerol) and incubated with nuclear extracts for 20 min in binding buffer containing 50 mM NaCl and 0.01% CHAPS using heating to 90°C for 5 min in TE buffer followed by slow cooling down at room temperature overnight. For competition experiments unlabeled oligonucleotides (CdxA 5’-G CATTTTATTACCCAGCTGCAGTGGTA-3’, MyoD 5’-CCCCCAACAGCTGTGGCCTGA-3’, Scramble 5’-AGCAACTTGGATAGTATAGGTCAAG ACCGCCACATT-3’, YY1 5’-CGCTCCCGGCCATC TTGGCGGCTGGT-3’) were used. Double-stranded oligonucleotides were separated from remaining single strand oligonucleotides on a 12% polyacrylamide gel.

Oligonucleotides probes

40 bp oligonucleotides rs4684847 5’-TTTAAATCTATCCTTAATTC[CT][C]/AACAATCCCGAAGAGATAAG-3’; rs7647481 5’-CAACTCCCCACTTATCC[A/G]/GTGATTAGCGATACCAGCCA-3’; rs17036342 5’-GCTCTCCCAAGAAATTTGAA[A/G]/CTCCAGA GTTGGAGACCA-3’; rs2881479 5’GCAAGACTCTGTC ATC[A/T]AAATTAAATATAAATAAAA-3’ with Cy5- (for EMSA) or biotin-label (for affinity chromatography) and with the respective SNP alleles at mid-position were synthesized (Eurofins, Ebersberg, Germany) and annealed with unmodified complementary oligonucleotides to obtain double stranded oligonucleotides, by
a rotator. We used 7 mg of nuclear extracts, which showed the most significant enrichment of proteins of interest. To reduce non-specific DNA binding, poly (di-dC) was subsequently added to the mixture and incubated further for 10 min. Competition with poly (di-dC) supports the specific protein binding we performed SYBR-green qPCR

Filter-aided sample preparation (FASP) and non-targeted LC–MS/MS

Salt eluted fractions were processed as described before (10,27) in an adaptation of the FASP approach (28) using Microcon devices YM-30 (Merck Millipore, Darmstadt, Germany). The LC–MS/MS analyses were performed as described previously on a LTQ-Orbitrap XL (Thermo Scientific, Dreieich, Germany) (29) with the following adjustments: A nano trap column was used (300 µm inner diameter × 5 mm, packed with Acclaim PepMap100 C18, 5 µm, 100 Å; LC Packings) before separation by reversed phase chromatography (PepMap, 25 cm, 75 µm ID, 2 µm/100 Å pore size, LC Packings, Thermo Scientific, Dreieich, Germany) operated on a RSLC (Ultimate 3000, Thermo Scientific, Dreieich, Germany) using a nonlinear 170 min LC-gradient from 5 to 31% of buffer B (98% acetonitrile) at 300 nl/min flow rate followed by a short gradient from 31 to 95% buffer B in 5 min and an equilibration for 15 min to starting conditions. From the MS pre-scan (acquired in profile mode), the 10 most abundant peptide ions were selected for fragmentation in the linear ion trap if they exceeded an intensity of at least 200 counts and if they were at least doubly charged. Dynamic exclusion was set to 30 s. During fragment analysis, a high-resolution (60 000 full-width half maximum) MS spectrum was acquired in the Orbitrap with a mass range from 300 to 1500 Da.

Protein identification and label-free relative quantification

The RAW files (Thermo Scientific, Dreieich, Germany) were analyzed using the Progenesis LC–MS software (version 4.0, Nonlinear Dynamics, Waters, Eschborn, Germany), as described previously (29,30), with the following changes: Spectra were searched using the search engine Mascot (Matrix Science, London, UK) against the Ensembl mouse database (Release 69; 50512 sequences). False discovery rates were stringently kept below 1% as calculated by a Mascot-integrated decoy database search using the percolator algorithm (cut-off score 13, significance threshold of P < 0.05). Peptide assignments were re-imported into Progenesis LC–MS. Normalized abundances of all unique peptides were summed up and allocated to the respective protein.

Allele-specific YY1 Chromatin Immunoprecipitation assay (ChIP)

For each experiment, 1 × 10^6 primary human preadipocytes as well as preadipocytes in vitro differentiated to adipocytes for 14 days were cultured in six wells of a six-well plate as previously described (6). ChIP experiments were performed using the ChIP-IT® Express Enzymatic Chromatin Immunoprecipitation Kit from Active Motif (La Hulpe, Belgium) according to the manufacturer’s protocol with slight modifications as described elsewhere (31). Briefly, after enzymatic digestion for 15 min, 10 mM EDTA was added and chromatin was sheared using the EpiShear™ Probe Sonicator (Active Motif, La Hulpe, Belgium; 20 pulses consisting of 20 s sonication followed by 30 s rest at 25% amplitude) in the same buffer. Chromatin was then incubated for 30 min with protein G magnetic beads (Active Motif, La Hulpe, Belgium) and 2 μg of rabbit polyclonal αYY1 antibody (sc-281x, Santa Cruz Biotechnology, Heidelberg, Germany). Incubation with 2 μg of rabbit IgG (sc-2027x) served as internal negative controls. The amount of precipitated DNA was evaluated by allele-specific quantitative PCR (AS-qPCR) using the Eppendorf Mastercycler (Eppendorf, Hamburg, Germany). To quantify allele-specific protein binding we performed SYBR-green qPCR (Maxima SYBR Green/ROX qPCR Master Mix, Thermo Scientific, Dreieich, Germany) using rs7647481 forward primer (5′-AAGATGTGGGGGCTTATGGA-3′) with the allele-specific reverse primers rs7647481A nonrisk (5′-GCTGGTCGAAACATGATG-3′) and G-risk (5′-CTGAGGGCTGAAACATGACG-3′), respectively (bold: SNP position, underlined: additional mutation). Allele-specific reverse primers were designed (TIB Molbiol, Berlin, Germany) with the respective rs7647481 allele-specific ChIP-chromatin Ct-values for both anti-YY1 and IgG experiments, the allele-specific ratio to the respective IgG control for each experiment. Pairs of anti-YY1 and IgG ChIP experiments were performed from fixed chromatin of preadipocytes and in vitro differentiated adipocytes from three donors, which were previously genotyped with a concordance rate of >99.5% using the MassARRAY system with iPLEX™ chemistry (Sequenom, Hamburg, Germany) as described (33).

siRNA knock down and PCR

SGBS cells were cultured in 6-well plates and transfected with 25 nM siRNA targeting YY1, RYBP or as control non-targeting (NT) siRNA (ON-TARGETplus human siRNA SMARTpool, Dharmacon, Freiburg, Germany)
using HiPerFect (Qiagen, Hilden, Germany) according to the manufacturer’s instructions. Seventy two hours after transfection, confluent cells were harvested using the RNeasy-MiniKit (Qiagen, Hilden, Germany) to extract total RNA and stored at –80°C. The high capacity cDNA Reverse Transcription kit (Applied Biosystems, Darmstadt, Germany) was used for transcription of 1 μg total RNA into cDNA according to the manufacturer’s instructions and stored at –20°C. RNA/cDNA concentration and purity (A260/A280) was assessed using NanoQuant Plate™ (TECAN, Männedorf, Switzerland). qPCR analysis to the housekeeping gene, GAPDH, was performed using the qPCR SYBR-Green ROX Mix (ABgene, Hamburg, Germany) to the manufacturer's instructions. Seventy two hours after transfection, the cells were washed once with PBS and lysed in 1× passive lysis buffer (Promega, Mannheim, Germany) on rocking for 20 min at room temperature. Firefly and renilla luciferase activity were measured using luminometer (Berthold, Pforzheim, Deutschland), respectively. The ratios of firefly luciferase expression to renilla luciferase expression were calculated and normalized to the thymidine kinase promoter control vector. All experiments were performed repeatedly as indicated. P-values were calculated using paired t-test.

**GO-term, signaling pathway and transcription factor/transcriptional coregulator co-citation analysis**

The Genomatix GePS-tool (Genomatix, Munich, Germany) was used to assess the enrichment of molecular function GO-terms (Supplementary Table S2) and signaling pathways (Supplementary Table S3) in the protein/gene data sets identified by LC-MS/MS, using all identified proteins. Next, To calculate the enrichment of transcriptional coregulators, we used the Genomatix GePS-tool to build a co-citation based network. All proteins annotated as co-factors (Supplementary Table S4) and the respective candidate transcription factors PRRX1, YY1 and NFI TC4 were used as input gene list to create networks using the settings co-citation level: *Function word level; co-citation filter: (i) Network generation: simple network; and additional interactions per gene: (ii) Genes with interactions are shown in the Figure 5A and Supplementary Figure S5A-C.

**Regression analysis of human adipose tissue samples**

The insulin-resistance measure HOMA-IR (homeostasis model assessment of insulin resistance) and mRNA expression levels were measured in a cohort comprising 30 obese (BMI > 30 kg/m²) otherwise healthy and 26 non-obese (BMI < 30 kg/m²) healthy women (34), all pre-menopausal and free of continuous medication and investigated in the morning after an overnight fast. Venous blood sample was obtained for measurements of glucose, insulin and for preparation of DNA. HOMA-IR was calculated by the formula \(\text{IP-}\text{Glucose (mmol/L) } \times (\text{FS-Insulin (microU/mL)}/22.5 (35). Following blood sampling abdominal subcutaneous adipose tissue biopsy was obtained by needle aspiration and adipose microarray analysis was performed exactly as described (34) using the Affymetrix GeneChip Array protocol with 1 μg of total adipose RNA from each subject (gene expression deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO; https://ncbi.nlm.nih.gov/geo/) and accessible using GEO series accession number GSE25402). Linear regression analyses were performed with R, version 3.0.2 (R: A Language and Environment for Statistical Computing: from the R Development Core Team of the R Foundation for Statistical Computing, Vienna, Austria, 2014, http://www.R-project.org/) to assess correlation of YY1, RYBP and PPARG adipose tissue mRNA levels with HOMA-IR (adjusted for age, age/BMI and without...
Statistical analysis

All results were expressed as mean ± SD. Student’s t-tests, Wilcoxon signed rank test and one-sample t-tests were used to compare two groups (two alleles/complex and non-complex SNP regions). All statistical analyses were performed using the Graph Pad Prism software v5.02 (GraphPad Software, CA, USA) or the Statistical Software SPSS v20.0 (IBM Corporation, NY, USA). The applied statistical methods for each experiment are given in the respective figure legend. Statistical differences of results were shown in *P < 0.05, **P < 0.01, ***P < 0.001.

RESULTS

Risk and nonrisk allele-specific protein–DNA interaction at the cis-regulatory variants of rs4684847 and rs7647481

At the PPARG locus, with a total of 23 non-coding variants in high linkage disequilibrium \((r^2 \geq 0.7\) with PPARG tagSNP rs1801282 (6)), six variants were predicted to be cis-regulatory by conserved binding site modularity analysis (Figure 1B). We showed previously that rs4684847 specifically overlaps with a homeobox TFBS and modulates regulation of PPARG expression (6). Further assessing proximity of the remaining five variants to marks of regulatory regions in adipocytes, we find rs7647481 to overlap with all histone modification regions (36) at both tested late stages (day 3 and day 9) of human adipogenesis (Figure 1C), and with DNase-seq regions (2) in differentiated primary human adipocytes (Supplementary Figure S1); supporting a contribution of rs7647481 to adipocyte specific PPARG2.
expression (Figure 1D). Both, rs7647481 and the previously reported rs4684847 (6) are in perfect linkage disequilibrium ($r^2=1.0$ in 1000 Genomes (37)), confirmed by sequencing in this study for all analysed tissues and cells (data not shown). To experimentally evaluate these cis-regulatory predictions, we further analysed allele-specific protein–DNA interactions at both predicted cis-regulatory variants and additionally at two variants predicted as non cis-regulatory (Figure 1D).

Electrophoretic mobility shift assays (EMSA) using nuclear extracts of mouse brown adipocytes, revealed specific protein–DNA interaction for both predicted cis-regulatory SNPs, not observed with the predicted non cis-regulatory SNPs (Figure 2A upper panel). Quantification of protein–DNA complexes confirmed allele-specific binding at both predicted cis-regulatory SNPs ($P=0.03$) in contrast to non cis-regulatory SNPs ($P=0.82$ for rs17036342 and $P=0.80$ for rs2881479, respectively) (Figure 2A lower panel). PPARG regulates adipogenesis and mature adipocyte metabolism. In EMSA experiments using nuclear protein extract from primary human preadipocytes, human SGBS cell strain preadipocytes and in vitro differentiated SGBS adipocytes, we find differential protein–DNA interaction patterns (Figure 2B), indicative for different transcription factors and regulatory protein complexes contributing to PPARG expression in different stages of adipogenesis. Allele-specific differential binding was observed consistently at predicted cis-regulatory variants, with increased binding at rs7647481A nonrisk and rs4684847C risk allele, respectively.

Identification of allele-specific binding proteins at predicted cis-regulatory variants by label-free quantitative proteomics

Next, we aimed to identify allele-specific binding proteins, i.e. regulators and coregulators at those variants by an unbiased, quantitative label-free protein–DNA proteomics approach (Figure 1A-5). To enrich allele-specific binding proteins for cis-regulatory variants, with in vitro differentiated SGBS adipocytes, we found differential protein–DNA interaction patterns (Figure 2B), indicative for different transcription factors and regulatory protein complexes contributing to PPARG expression in different stages of adipogenesis. Allele-specific differential binding was observed consistently at predicted cis-regulatory variants, with increased binding at rs7647481A nonrisk and rs4684847C risk allele, respectively.

Figure 3. Label-free quantitative LC–MS/MS proteomics identified risk versus nonrisk allele-specific binding proteins at predicted cis-regulatory and non cis-regulatory SNPs. Volcano plots for the indicated variants illustrate the distribution of risk (blue) and nonrisk (green) allele-specific binding proteins identified by LC–MS/MS (results from 200 mM NaCl eluates illustrated by EMSA in Figure 2C and Supplementary Figure S2, results from 300 mM eluates see Supplementary Figure S3) at predicted cis-regulatory (red) and non cis-regulatory SNPs (grey). Proteins with significant ($P \leq 0.05$, red line) allele-specific differential binding (allelic ratio $\leq 0.5$ or $\geq 2$) at the risk allele = blue dots, nonrisk allele = green dots; with no significant allele-specific binding = grey dots; n of proteins per quartile = italic numbers. Arrows highlight proteins from Table 1 annotated as transcription factors with fold change $\geq 2$, $P \leq 0.01$, number of identified peptides $\geq 2$. Mean protein levels (log$_2$ ratio of indicated alleles) and $P$-value from unpaired t-test of three independent experiments.
and significant differential binding (allelic fold change intensity after affinity-chromatography enrichment at rs7647481A nonrisk and rs4684847C risk allele (Figures 2C, 200 and 300 mM elution, respectively) and significant differential binding (allelic fold change ≥ 2.0, P-value ≤ 0.01, illustrated in Figure 3 and Supplementary Figure S3) are shown.

Focusing on transcription factors identified at the novel cis-regulatory variant rs7647481, label-free proteomics found the nonrisk allele binding transcription factors YY1 with a 6.6-fold (P = 2.94 × 10^{-3}) and NFATC4 with a 2.6-fold (P = 0.01) allelic fold-change (Table 1, Figure 3). For rs4684847, previously shown to abrogate a type 2 diabetes-specific homeobox TFBS and to infer with PRRX1 homeobox protein binding (6), proteomics found increased risk allele binding of PRRX1 and ILF3 (2.6-fold, P = 0.01 and 4.2-fold, P = 0.01, respectively, Table 1, Supplementary Figure S3). Subsequently, we assessed the enrichment of canonical signaling pathways using the GePS-tool (Genomatix) within the set of all identified allele-specific binding proteins and the occurrence of candidate transcription factors in the identified pathways (Supplementary Table S3). Notably, the only transcription factor identified in our screen included in significantly enriched signaling pathways was YY1 (P < 0.05, Fisher’s exact test, E2f transcription factor network, p53 pathway, Pre2 complex sets long-term gene silencing through modification of histone tails, and signaling events mediated by HDAC Class I).

YY1 drives transcriptional activity at the rs7647481A nonrisk allele of the PPARG locus

The common rs7647481G risk allele abrogates the core of a YY1 consensus TFBS, confirming the mass spectrometric identifications and GO-term analyses (Figure 4A). The allele-specific protein–DNA interaction at the rs7647481-adjacent region was efficiently blocked in competition and supershift EMSA experiments by 33-fold molar excess of unlabeled YY1 consensus binding sequence or by pre-incubation with a YY1 specific antibody (Figure 4B left panel), while protein binding was not affected at all other tested SNP-adjacent regions, including the cis-regulatory variant rs4684847 (Figure 4B). Competition with unspecific competitor oligonucleotides (consensus MyoD myogenic regulatory factors, consensus CdxA chicken homeodomain protein, and scrambled control sequence) did not affect the allele-specific protein binding (Supplementary Figure S4A), further confirming the specificity of YY1 binding at the rs7647481-adjacent region. PPARG expression is essential for preadipocyte differentiation and metabolic function of mature adipocytes. By YY1 competition and supershift experiments we confirmed YY1 binding at the rs7647481A nonrisk allele in both, 3T3-L1 preadipocytes and adipocytes (Supplementary Figure S4B). In reporter gene assays with luciferase constructs of all tested variants transfected into 293T cells, overexpression of the transcription factor YY1 revealed a significant higher transcriptional activity at the rs7647481A nonrisk allele as compared to the risk allele (P = 0.003, Figure 4C). Although we found an overall increase of all tested reporters, no allele-specific effects of the cis-regulatory variant rs4684847 and both non cis-regulatory variants were observed. The rs7647481A nonrisk variant also increased transcriptional activity in different cell types significantly, i.e. by 1.2-fold in 293T cells, 1.4-fold in INS1 β-cells, 2.7-fold in C2C12 myoblasts, 2.2-fold in C2C12 myocytes, 1.3-fold in Huh7 hepatocytes (P < 0.01, Figure 4D) and 1.5-fold in 3T3-L1 adipocytes (6).

Next, we aimed to confirm in vivo allele-specific binding of YY1 at the rs7647481A nonrisk allele, as results from reporter assays are limited by missing chromosome context or length of analysed sequences. We performed chromatin immunoprecipitation (ChIP) combined with allele-specific quantitative PCR (AS-qPCR) to test for allelic imbalance in primary human adipose tissue cells heterozygous for rs7647481G/A. ChIP experiments were performed using a YY1-specific antibody to pull down cross-linked YY1-protein / chromatin complexes from primary human preadipocyte and adipocyte cells of three donors (Figure 4E). In both, input chromatin and immuno-precipitated chromatin, the nonrisk and risk allele, respectively, were analyzed using AS-qPCR. In adipocytes we observed a significant mean 31.8-fold (P = 0.02, Wilcoxon signed rank test) increased binding of YY1 at the rs7647481A nonrisk allele as compared to the risk allele in chromatin immunoprecipitated with anti-YY1. In preadipocytes, we observed a large variation of YY1 binding at the nonrisk allele but no significant allele-specific binding. No allele-specific differences were observed when using an IgG control antibody. Overall, YY1 supershift (Figure 4B, Supplementary Figure S4B) and ChIP (Figure 4E) experiments in preadipocytes and adipocytes establish binding of the transcription factor YY1 at the rs7647481A nonrisk allele, and support its role in transcriptional regulation of the PPARG gene in both stages of adipogenesis.

Identification of transcription factor related transcriptional coregulators

Metabolic homeostasis is largely regulated at the transcriptional level through the coordinated interaction between transcription factors, coregulators, and the basal transcriptional machinery (11). Our pull down of protein–DNA binding complexes offers the opportunity to identify functional protein–DNA interactions as we found both, transcription factors (Table 1) and co-eluting transcriptional coregulators (Supplementary Table S4). We assessed relevant protein-protein interaction using literature co-citation.

Table 1. Allele-specific differentially binding transcription factors

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene symbol</th>
<th>Allelic ratio</th>
<th>Allelic fold change</th>
<th>P-value (fold change)</th>
<th>Quantified peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7647481</td>
<td>YY1</td>
<td>A/G</td>
<td>6.6</td>
<td>2.94 × 10^{-3}</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>NFATC4</td>
<td></td>
<td>2.6</td>
<td>0.01</td>
<td>2</td>
</tr>
<tr>
<td>rs4684847</td>
<td>PRRX1</td>
<td>C/T</td>
<td>2.6</td>
<td>0.01</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>ILF3</td>
<td></td>
<td>4.2</td>
<td>0.01</td>
<td>4</td>
</tr>
</tbody>
</table>

Proteins annotated as transcription factors and identified by LC–MS/MS in the fractions with the highest allelic protein–DNA binding EMSA signal intensity after affinity-chromatography enrichment at rs7647481A nonrisk and rs4684847C risk allele (Figures 2C, 200 and 300 mM elution, respectively) and significant differential binding (allelic fold change ≥ 2.0, P-value ≤ 0.01, illustrated in Figure 3 and Supplementary Figure S3) are shown.
Figure 4. rs7647481A nonrisk allele-specific binding and transcriptional activity of the transcription factor YY1 inferred from proteomics analysis. (A) The rs7647481G risk allele abrogates the core of a YY1 consensus binding site (Matbase Matrix Library 9.1, Genomatix, Munich, Germany). (B) Competition and supershift EMSA experiments using risk (R) and nonrisk (NR) allele-specific Cy5-labeled probes of the predicted cis-regulatory (red) and non cis-regulatory (grey) variants reveal a specific binding of YY1 at the rs7647481A nonrisk allele. Competition (comp.) assays using 33-fold excess of unlabeled YY1 probe and supershift assays by adding anti-YY1 (αYY1) or IgG isotype control antibody, respectively. (C) Reporter assays in 293T cells with constructs harbouring the risk and nonrisk allele of predicted cis-regulatory (red) and non cis-regulatory (grey) variants reveal allele-specific activation from the rs7647481A nonrisk allele upon YY1 overexpression. Mean ± SD from five experiments. (D) rs7647481A nonrisk allele-specific activation of reporter gene activity in 293T-cells, INS-1 β-cells, C2C12 cells (undifferentiated myoblasts, differentiated myocytes) and Huh7 hepatocytes assessing the effect of endogenous transcriptional regulators. Reporter assays with luciferase constructs containing the respective allele at midposition as indicated; for each cell line the TK-promoter control vector was co-transfected separately and set to one. Mean ± SD from seven experiments. (E) Increased in vivo YY1 binding at the rs7647481A nonrisk allele. The result shows allele-specific binding at the A-nonrisk/G-risk allele, determined by ChIP experiments in primary human adipose tissue cells, preadipocytes and in vitro differentiated adipocytes, heterozygous for rs7647481G/A using αYY1 or IgG isotype control, respectively, followed by allele-specific qPCR detecting the rs7647481A nonrisk and G risk allele for each ChIP experiment (see also Materials and Methods). Mean ± SD from ChIP experiments using chromatin–DNA from three donors, P-values from Wilcoxon's signed rank test.
of the allele-specific binding transcription factors YY1, \textit{NFATC4}, \textit{PRRX1} and \textit{ILF3} (Table 1) with all identified coregulators for prioritization of in-depth proof of concept experiments. Note, that considering the second order binding of cofactors to transcription factors binding at DNA, we included all cofactors identified by proteomics regardless of fold-change (Supplementary Table S4). In contrast to \textit{NFATC4}, \textit{PRRX1} or \textit{ILF3}, we found a significant enrichment of reported transcriptional coregulators co-cited with YY1 ($P = 1.56 \times 10^{-3}$, fishers exact test, Methods), i.e. RING1 and YY1 binding protein (RYBP), YY1 associated factor 2 (YAF2), prohibitin (PHB), nucleophosmin (NPM1), host cell factor C1 (HCFC1), metastasis associated 1 family (MTA2), DEK oncogene (DEK), and high mobility group box 2 (HMG2B). Additionally, visualizing annotated gene-gene interactions of the transcription factor YY1 with all proteomics-identified proteins annotated as cofactors reveals a network connecting YY1 with RYBP, NPM1, YAF2, MTA2, HCFC1 and metadherin (MTDH) (Figure 5A). Co-localization of YY1 and RYBP has been reported previously by using immunofluorescence staining and co-immunoprecipitation experiments (38). \textit{NFATC4} and \textit{ILF3} were found connected with the cofactor calreticulin (CALR) and DEAH-box helicase 9 (DHX9), respectively, none of the identified cofactors was connected with \textit{PRRX1} (Supplementary Figure S5A–C).

For the novel cofactors which were identified by our unbiased proteomics but were not co-cited with YY1 or NFATC4 at rs7647481A nonrisk and \textit{PRRX1} or \textit{ILF3} at rs4684847C risk allele (Supplementary Table S4) further experimental proof will be necessary to support protein-protein interaction. Here, we focused on experimental proof for the YY1 / RYBP interaction.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure5}
\caption{Interaction network analysis of YY1 with cofactors infers RYBP contribution to nonrisk allele-specific effect on insulin-resistance. (A) Interaction network of the YY1 transcription factor identified at the rs7647481A nonrisk allele with all transcriptional coregulators identified in the same label-free proteomics association. Associations by co-citation (dotted lines) or expert curation (lines) from GePS tool analysis (Genomatix, see Materials and Methods). Proteins with direct interaction to the transcription factor YY1 (green dotted line) and with positive correlation of adipose mRNA levels to insulin-sensitivity (green line) are shown (Table 2). (B and C) YY1 and RYBP (B). \textit{PPARG1} and \textit{PPARG2} (C) mRNA expression levels measured by qPCR, standardized to \textit{GAPDH}, in SGBS preadipocytes treated with different siRNAs for 72 h labeled as siYY1, siRYBP or siYY1+siRYBP/siNT (non-targeting control). Mean ± SD from five experiments. $P$-values from one sample t-test. (D) Impact of nonrisk and risk allele identified proteins on the \textit{PPARG} locus phenotype insulin-resistance. The rs4684847C risk allele binds the \textit{PPARG} suppressor \textit{PRRX1}; resulting in insulin-resistance (6).

\end{figure}

\textbf{Allele-specific correlation of transcription factor and cofactor expression levels in adipose tissue with insulin resistance}

Next, we assessed if the risk and nonrisk allele-specific proteomics findings can be related to disease pathophysiology. The minor nonrisk allele of the \textit{PPARG} locus (tagSNP rs1801282 Pro12Ala) was repeatedly associated with improved insulin-sensitivity in diverse studies. Thus, allele-specific \textit{PPARG} expression may contribute to this phenotype, given the essential role of \textit{PPARG} to maintain insulin-sensitivity. In fact, testing allele-specific \textit{PPARG} mRNA levels in adipose tissue samples, we observed a confirmative age- and age/BMI-independent negative correlation of total \textit{PPARG} mRNA levels with the insulin resistance measure HOMA-IR in nonrisk allele carriers ($\beta = -6.25$, $P = 2.18 \times 10^{-4}$; $\beta = -3.20$, $P = 0.05$, respectively) as compared to risk allele carriers ($\beta = 0.28$, $P = 0.89$; $\beta = 0.10$; $P = 0.92$, respectively, Table 2). A coordinated regulation of \textit{PPARG} expression by YY1 and co-identified coregulators at the rs7647481A nonrisk allele may contribute to phenotypes—such as insulin-resistance—associated with the \textit{PPARG} locus. Here, for adipose tissue mRNA expression levels of the coregulator RYBP, identified by proteomics and reported to interact with the YY1 transcription factor (39), we found a negative age- and age/BMI-independent correlation with HOMA-IR in individuals carrying the nonrisk ($\beta = -5.71$, $P = 1.15 \times 10^{-3}$; $\beta = -3.38$, $P = 7.04 \times 10^{-5}$; respectively) as compared to risk ($\beta = 0.57$, $P = 0.67$; $\beta = 0.16$, $P = 0.85$, respectively) allele. For none of the other proteins co-cited with YY1 or \textit{NFATC4} (Supplementary Table S5) an allele-dependent and BMI-independent correlation was observed. We found a BMI-dependent correlation of \textit{PHB}, \textit{CALR} and \textit{NPM1} expres-
bation (also found at rs4684847C risk allele with PRRX1 and ILF3), requiring further analysis in the future. Taking into account the impact of BMI on insulin-sensitivity, our data support the literature-mining based initial focus on YY1 and RYBP. While we found no significant correlation of YY1 mRNA levels with HOMA-IR in the small available data set, a confirmative direction of \( \beta \)-values was observed (Table 2). Finally, we also found a positive correlation for adipose mRNA expression levels of both, YY1 and RYBP, with total mRNA levels of the insulin-sensitizing transcription factor PPARG from both alleles (Supplementary Table S5). YY1 and RYBP can function as repressor or activator depending on genomic context and availability of interaction partners (38,40). Assessing the effect on endogenous mRNA expression levels in SGBS preadipocytes, we found that knockdown of YY1 and its cofactor RYBP alone (Figure 5B, 63% or 60% knockdown efficiency, \( P = 7 \times 10^{-3} \) or 4 \( \times 10^{-4} \), respectively) resulted in a slight but not significant 1.6-fold increase of PPARG2 expression (Figure 5C) suggesting inhibitor-activity. However, in 293T cells YY1 overexpression significantly increased rs7647481A nonrisk allele reporter activity (Figure 4C) suggesting activator-activity, possibly by recruitment of endogenously expressed RYBP (Supplementary Figure S5D). Therefore, we performed simultaneous knockdown of YY1 and RYBP (Figure 5B, 68% or 73% knockdown efficiency, \( P = 3 \times 10^{-3} / 2 \times 10^{-3} \), respectively) in the PPARG1 and PPARG2 expressing adipose tissue cell line SGBS. We found a significant 2-fold reduction of endogenous mRNA expression levels for the insulin-sensitizing PPARG2 isoform (\( P = 0.027 \), Figure 5C) suggesting that YY1 and RYBP jointly activate PPARG2 expression and supporting the importance to account for both, transcription factors and related cofactors. How further cofactors identified by our unbiased proteomics approach (Supplementary Table S4) contribute to activation or inhibition of PPARG gene expression requires future experiments. Overall, our findings suggest that PPARG expression, and thereby insulin sensitivity, may be increased by rs7647481A nonrisk allele-specific binding of the transcription factor YY1 and its cofactor RYBP (Figure 5D).

**DISCUSSION**

GWAS revealed numerous risk loci associated with common traits (1). Advances of the ENCODE project and novel bioinformatics approaches facilitate identification of cis-regulatory, potentially diseases-causing variants within complex loci (2–7). However, the spatial and temporal varied expression pattern of transcription factors and coregulators supports the need to consider cell-type specific open chromatin data (2.41–45) to prioritize candidate cis-regulatory variants. In our study, combining computational TFBS modularity analysis (6) and functional cell type- and differentiation-specific data from human adipocytes (36) enabled identification of a second cis-regulatory variant at the PPARG locus, additional to a previously reported (6). Our unbiased quantitative label-free proteomics approach came up with significantly more allele-specific binding proteins at predicted cis-regulatory as compared to non cis-regulatory variants, supporting the integrative framework predictions. Moreover, we find both, allele-specific transcription factors and cofactors which may contribute to the rs7647481A nonrisk allele-specific association of the PPARG locus with insulin sensitivity.

Identifying the allele-dependent binding proteins is an essential step to further decipher the molecular mechanisms underlying genetic associations with disease pathophysiology and ultimately to define personalized interventions. However, identification of allele-specific binding proteins by TFBS matrix overlap or ChIP-seq faces limitations, such as availability of TFBS matrix annotation and the complexity of TFBS modularity. So far, few proteomic studies have provided successful transcription factor identification at cis-regulatory variants (12,13,46). Here, taking into account the importance of TFBS modularity for specific protein-DNA interaction (6,47,48), we assessed protein binding directly at the genomic regions of interest, instead of e.g. concatenated oligonucleotides (13). Peptide intensity-based label-free quantification across samples was previously benchmarked against SILAC labeling and compared well with respect to reproducibility, sensitivity, and robustness (14). In contrast to metabolic (13,46) or chemical labeling strategies (12) our label-free approach does not limit future application to any tissue or cell type, including primary human tissue. Our proteomics workflow on eluted fractions containing allele-specific binding-proteins allows identification of allele-specific binding transcription factors. Moreover, our workflow enables for the first time the identification of transcriptional coregulators related to allele-specific gene regulation, which are among the lowest abundant proteins in cells (42,49).

The PPARG locus is robustly associated with type 2 diabetes and insulin-sensitivity (16–18,50,51) and mRNA levels of the insulin-sensitizing PPARG are increased in non-risk allele carriers (6). At the here identified rs7647481A nonrisk allele, proteomic analysis identified allele specific binding of two transcription factors, YY1 and NFATC4. YY1 was reported to regulate metabolic, diabetes-related phenotypes in skeletal muscle, liver and adipocytes (52–56). NFATC4 contributes to regulation of PPARG and mouse adipocyte differentiation by direct binding at the PPARG2 promoter (57), YY1 indirectly by interaction with C/EBP\( \alpha \) (56). We focused on the allele-specific activity of human YY1 and found in vivo allele-specific binding at the rs7647481A nonrisk allele in the human PPARG promoter by ChIP experiments combined with assessment of allelic imbalance in heterozygous primary human adipose tissue cells, supporting the biological relevance of the proteins identified by our label-free proteomics approach. At the rs4684847C risk allele, we previously reported an overlapping homeobox TFBS, risk allele specific binding of the transcriptional inhibitor PRRX1 repressing PPARG expression and promoting insulin-resistance (6). Our unbiased proteomics confirmed risk allele binding of PRRX1 and also found ILF3, a transcription factor so far not related to adipose PPARG gene regulation or diabetes. The regulation of PPARG expression is complex, and future experiments are needed to support contribution of further factors, such as NFATC4 and ILF3 at rs7647481 and rs4684847, respectively, to allele-specific PPARG expression in human adipose tissue, but also for possible interactions of PRRX1 with ILF3 or NFATC4 with YY1.
Table 2. Risk and nonrisk allele specific correlation of adipose tissue PPARG, YY1 and RYBP mRNA expression levels (log transformed) with the type 2 diabetes trait insulin-resistance

<table>
<thead>
<tr>
<th>Gene</th>
<th>Allele</th>
<th>Adj</th>
<th>HOMA IR</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPARG</td>
<td>All</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a.b</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Nonrisk</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a.b</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Risk</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a.b</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>YY1</td>
<td>All</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a.b</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Nonrisk</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a.b</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Risk</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a.b</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>RYBP</td>
<td>All</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a.b</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Nonrisk</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a.b</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Risk</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>a.b</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Gene expression was measured in adipose tissue from a lean/obese patient cohort (38 subjects, mean ± SD 24.2 ± 9.1 kg/m²). rs7647481 and rs4684847 risk allele and nonrisk allele genotypes were determined by Sequenom-assay. Nonrisk: subjects heterozygous or homozygous (n = 18) for the rs7647481A (YY1/RYBP binding) and rs4684847T nonrisk allele; risk: subjects homozygous (n = 20) for the rs7647481G and rs4684847C risk allele. P-values and \( \beta \)-estimates from linear regression analysis of total PPARG mRNA levels (from microarray data measuring exons shared by both PPARG1 and PPARG2), YY1 and RYBP mRNA expression levels with insulin-resistance measure HOMA-IR (homeostasis model assessment of insulin resistance) are shown. Adj = correlations without adjustment (—), age (a) or age and BMI adjusted (a.b).

In addition to transcription factors, the high sensitivity of the applied proteomics approach enabled identification of several cofactors which may be involved in modulating transcriptional regulation of PPARG, by yet to be characterized mechanisms in future studies. For proof-of-concept that our allele-specific proteomics can find both, allele-specific transcription factors and cofactors, we selected the transcription factor YY1 with the highest allele-specific binding and the co-citation enriched RYBP cofactor at rs7647481. The coregulator RYBP—known to interact with YY1 and here co-identified at the rs7647481 variant—was recently shown to associate with skeletal myogenesis (38), in addition to its function as transcription repressor in cancer (58,59), embryogenesis (39) and central nervous system development (60). While YY1 has been reported previously to promote adipocyte differentiation (61), we found that the combined action of the transcription factor YY1 and the cofactor RYBP is necessary for full activation of PPARG2 isoform expression in an adipocyte cell line. PPARG2 has been reported to be crucial for maintaining insulin-sensitivity (20). In human samples we demonstrated a nonrisk allele-specific association of RYBP expression levels with the PPARG nonrisk allele associated phenotype insulin-sensitivity. Thus, in carriers of the protective PPARG allele, binding of YY1 and RYBP at the rs7647481A nonrisk allele may increase PPARG2 expression and thereby insulin sensitivity (Figure 5D, upper panel); and in PPARG risk allele carriers the previously reported rs4684847C risk allele binding of PRRX1 inhibits PPARG expression thereby promoting insulin-resistance (6) (Figure 5D, lower panel).

In conclusion, our results demonstrate the importance to consider allele-specific protein complexes for delineation of the molecular mechanisms affected by cis-regulatory variants. We present an approach enabling an unbiased identification of allele-specific protein-DNA interactions including both, transcription factors and transcriptional cofactors. Additionally, we provide data supporting that different cis-regulatory variants at the PPARG locus may contribute to disease pathophysiology (Figure 5D), in line with the ‘multiple enhancer hypothesis’ suggesting multiple regulatory SNPs per LD block (62). Integrative approaches combining computational and cell type specific histone / chromatin mark based cis-regulatory prediction in combination with highly sensitive proteome-wide identification of allele-specific binding proteins can help to clarify the phenotypic effects of inherited and somatic genetic variability.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank M. Arnold and C. Fuchs for statistics, G. Lebouille for advice with allele-specific primer design, E. Hofmair, C. Herrmann, J. Behler, S. Becker and M. Hubsberger for technical assistance. We acknowledge the
REFERENCES


